

Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study

Valentina Bellemo, Zhan W Lim, Gilbert Lim, Quang D Nguyen, Yuchen Xie, Michelle Y T Yip, Haslina Hamzah, Jinyi Ho, Xin Q Lee, Wynne Hsu, Mong L Lee, Lillian Musonda, Manju Chandran, Grace Chipalo-Mutati, Mulenga Muma, Gavin S W Tan, Sobha Sivaprasad*, Geeta Menon*, Tien Y Wong*, Daniel S W Ting*



Summary

Background Radical measures are required to identify and reduce blindness due to diabetes to achieve the Sustainable Development Goals by 2030. Therefore, we evaluated the accuracy of an artificial intelligence (AI) model using deep learning in a population-based diabetic retinopathy screening programme in Zambia, a lower-middle-income country.

Methods We adopted an ensemble AI model consisting of a combination of two convolutional neural networks (an adapted VGGNet architecture and a residual neural network architecture) for classifying retinal colour fundus images. We trained our model on 76 370 retinal fundus images from 13 099 patients with diabetes who had participated in the Singapore Integrated Diabetic Retinopathy Program, between 2010 and 2013, which has been published previously. In this clinical validation study, we included all patients with a diagnosis of diabetes that attended a mobile screening unit in five urban centres in the Copperbelt province of Zambia from Feb 1 to June 31, 2012. In our model, referable diabetic retinopathy was defined as moderate non-proliferative diabetic retinopathy or worse, diabetic macular oedema, and ungradable images. Vision-threatening diabetic retinopathy comprised severe non-proliferative and proliferative diabetic retinopathy. We calculated the area under the curve (AUC), sensitivity, and specificity for referable diabetic retinopathy, and sensitivities of vision-threatening diabetic retinopathy and diabetic macular oedema compared with the grading by retinal specialists. We did a multivariate analysis for systemic risk factors and referable diabetic retinopathy between AI and human graders.

Findings A total of 4504 retinal fundus images from 3093 eyes of 1574 Zambians with diabetes were prospectively recruited. Referable diabetic retinopathy was found in 697 (22.5%) eyes, vision-threatening diabetic retinopathy in 171 (5.5%) eyes, and diabetic macular oedema in 249 (8.1%) eyes. The AUC of the AI system for referable diabetic retinopathy was 0.973 (95% CI 0.969–0.978), with corresponding sensitivity of 92.25% (90.10–94.12) and specificity of 89.04% (87.85–90.28). Vision-threatening diabetic retinopathy sensitivity was 99.42% (99.15–99.68) and diabetic macular oedema sensitivity was 97.19% (96.61–97.77). The AI model and human graders showed similar outcomes in referable diabetic retinopathy prevalence detection and systemic risk factors associations. Both the AI model and human graders identified longer duration of diabetes, higher level of glycated haemoglobin, and increased systolic blood pressure as risk factors associated with referable diabetic retinopathy.

Interpretation An AI system shows clinically acceptable performance in detecting referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema in population-based diabetic retinopathy screening. This shows the potential application and adoption of such AI technology in an under-resourced African population to reduce the incidence of preventable blindness, even when the model is trained in a different population.

Funding National Medical Research Council Health Service Research Grant, Large Collaborative Grant, Ministry of Health, Singapore; the SingHealth Foundation; and the Tanoto Foundation.

Copyright © 2019 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

Zambia is a lower-middle-income developing country located in southern Africa with a population of around 17 million people, which is expected to reach 20 million people in 2020. According to the International Monetary Fund, it has been projected to be ranked 159th (out of 194 countries) for gross domestic product per capita in 2018,¹ highlighting the need for cost-effective solutions

for Zambia to achieve their Sustainable Development Goals by 2030. As in all sub-Saharan African countries, Zambia's national priorities are focused on ways to improve life expectancy. The region, however, faces a critical shortage of health facilities and practitioners and alternate routes of public health delivery are required.²

Although available resources in low-income and middle-income countries have been diverted to tackling

Lancet Digital Health 2019;
1: e35–44

See [Comment](#) page e6

*Contributed equally

Singapore Eye Research Institute, Singapore National Eye Centre, Singapore (V Bellemo MSc, Q D Nguyen BEng, Y Xie MScPH, H Hamzah BSc, J Ho DFST, X Q Lee BSc (Hons), Gavin S W Tan MD, T Y Wong MD, D S W Ting MD); School of Computing, National University of Singapore, Singapore (Z W Lim PhD, G Lim PhD, W Hsu PhD, M L Lee PhD); Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore (M Y T Yip BA, G S W Tan, T Y Wong, D S W Ting); Eye Department, Kitwe Central Eye Hospital, Kitwe, Zambia (L Musonda MD); Department of Ophthalmology, Frimley Park Hospital NHS Foundation Trust, Camberley, UK (M Chandran FRCOphth, G Menon FRCOphth); Ophthalmology Department, Lusaka University Teaching Hospital, Lusaka, Zambia (G Chipalo-Mutati FCOphth (ECSA)); Ministry of Health, Lusaka, Zambia (M Muma FCOphth (ECSA)); and NIHR Moorfields Biomedical Research Centre, London, UK (S Sivaprasad FRCOphth)

Correspondence to: Assistant Prof Daniel Shu Wei Ting, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 168751 daniel.ting.s.w@singhealth.com.sg

Research in context

Evidence before this study

We evaluated the current state of diabetic retinopathy screening programmes by searching Scopus, Web of Science, PubMed, MEDLINE, and Embase for studies published in English from inception up to March 10, 2019, using the keywords “diabetes”, “diabetic retinopathy (DR) screening”, “fundus photographs”, “automated diabetic retinopathy system”, “machine learning”, “artificial intelligence (AI)”, and “deep learning”. We found many papers reporting robust diagnostic performance in detecting diabetic retinopathy, mostly from high-income countries (the USA, Singapore, Australia, and the UK) or low-income or middle-income countries with vast technical expertise (eg, China and India). However, none of the studies reported the use of artificial intelligence (AI) in under-resourced countries such as those in Africa. Zambia is a developing state in which a quarter of the population have eye diseases and visual impairment. Millions of Zambians have little access to appropriate care facilities and have a shortage of health practitioners. Therefore, the application of AI as an alternative screening tool for diabetic retinopathy could be of great benefit to the African population with diabetes.

Added value of this study

Our study shows an alternative clinically effective screening tool for diabetic retinopathy that uses deep learning to detect

referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema in an under-resourced African population with diabetes, with the data prospectively collected in Zambia. The diabetic retinopathy grading generated by the AI system could also accurately identify the exact same systemic risk factors as human graders.

Implications of all the available evidence

If deployed appropriately, AI using deep learning could be a clinically effective tool to detect vision-threatening diabetic retinopathy, which requires referral, in under-resourced countries with severe worker and infrastructure shortages. Additionally, this AI system could help to detect the systemic risk factors related to diabetic retinopathy, but with the ability to interpret the data more quickly than human graders. The incorporation of such AI systems into health care in under-resourced countries aligns with the VISION 2020 strategy to improve access and increase quality of eye-care services. Future research is beneficial to assess the generalisability, long-term outcomes, and cost-effectiveness of the AI-assisted diabetic retinopathy screening model, aiming to develop and maintain sustainable national eye-care programmes to prevent diabetic retinopathy-related blindness among the African population.

communicable diseases, these nations continue to face substantial public health and economic burden due to diabetes and its complications.² By 2040, 600 million people worldwide will have diabetes, of whom a third are expected to have diabetic retinopathy.^{2,3} Specifically in Africa, the prevalence of diabetes ranges from 6.3% to 15.4%,^{4,6} with 30.2–31.6% of diabetics with any diabetic retinopathy, 0.9–1.3% with proliferative diabetic retinopathy, and 1.2–4.5% with any maculopathy.³ Unfortunately, sub-Saharan Africa has the lowest average number of ophthalmologists per million population globally, with fewer than three ophthalmologists per million population, contrasting with approximately 80 ophthalmologists per million population in high-income countries.⁷ This shortage of ophthalmologists represents a substantial challenge to the health-care systems of under-resourced regions. Hence, it is essential to establish quick and cost-effective measures to identify and manage these morbidities with less strain on human resources.⁸

Diabetic retinopathy, a common diabetes-related microvascular complication, is the leading cause of blindness among working-age populations.^{9–11} A pooled analysis¹² of 35 studies found that the overall prevalence of the condition was 34.6% for any diabetic retinopathy, 10.2% for vision-threatening diabetic retinopathy, and 6.8% for diabetic macular oedema. Evidence from the few published reports from sub-Saharan countries indicates the prevalence of diabetic retinopathy ranges

from 7% to 63% and an increase in the proportion of blindness attributable to diabetic retinopathy is expected over the next 10 years.^{13,14} with one study showing up to 61 cases per 1000 people aged 50 years and older.¹⁵ A 5-year cohort study of African people with diabetes found that progression from no status of diabetic retinopathy to vision-threatening diabetic retinopathy was approximately five times higher in this population compared with European studies, and progression from background diabetic retinopathy was approximately three times higher.¹⁶ Moreover, a high prevalence of vision-threatening diabetic retinopathy has been reported in Zambia,¹⁷ close to the upper range of recent estimates.

The lack of workers, infrastructure, and public awareness are the key challenges that need to be tackled immediately, with the aim to prevent blindness in such countries. Over the past few years, artificial intelligence (AI) using deep learning and transfer learning algorithms has achieved excellent diagnostic performance in detecting major medical conditions,^{18,19} including various retinal eye diseases such as diabetic retinopathy,^{20–25} glaucoma,²⁶ age-related macular degeneration,^{27,28} retinopathy of prematurity,²⁹ and cardiovascular risks³⁰ from retinal fundus images. The diagnostic accuracy of AI technology in identifying these conditions has been shown to be comparable to the grading of diabetic retinopathy by retinal specialists on hospital-based retinal images. AI systems offer substantial benefits for health care by reducing the reliance on manual work and providing

savings in costs and resources, and this potential should be incorporated in screening programmes that are currently not widely implemented or routinely practised.

Here, we evaluated the diagnostic performance of an AI system using deep learning in detecting referable or vision-threatening diabetic retinopathy and diabetic macular oedema among an African population with diabetes. We also examined the associations between diabetic retinopathy detected by the AI model and systemic risk factors, hypothesising that the AI model would perform as accurately as human graders in identifying the systemic risk factors related to diabetic retinopathy.

Methods

Study population dataset

This clinical validation study was done using a diabetic retinopathy screening cohort of the Community Eye Service Programme, established by the Kitwe Central Hospital Eye Unit in partnership with Konkola Copper Mines (Chingola, Zambia), and Frimley Park Hospital Eye Department (Frimley, UK). This was the first mobile diabetic retinopathy screening programme in Zambia,¹⁷ aiming to improve public awareness among the diabetic population and to refer patients with vision-threatening diabetic retinopathy promptly to the tertiary eye-care services for early intervention. Patients with type 1 or type 2 diabetes were identified either through diabetes or pharmacy registries and were invited for screening either via local billboard advertising, radio and TV broadcasts, or within the church congregations. Almost 70% of the patients' diabetes type were not known.

During the screening, two-field (macular centred and retinal centred), colour, non-stereo, 45° retinal fundus photographs were taken for each eye, using the Digital Retinopathy System (CentreVue; Fremont, CA, USA) fundus camera. The retinal images (sometimes more than two per eye in cases of low quality or uncertain diagnosis) were captured with JPEG compression format, with dimension 2592 × 1944 pixels. The dataset also provided patient demographics and risk factors (eg, age, sex, diabetes duration, diabetes type, glycated haemoglobin [HbA_{1c}], body-mass index, systolic blood pressure, and diastolic blood pressure). The diabetic retinopathy screening programme¹⁷ was previously approved by Tropical Diseases Research Centre (Ndola, Zambia). Subsequently, this specific study was also approved by the Centralized Institutional Review Board of SingHealth, Singapore (protocol number SHF/FG648S/2015) and conducted in accordance with the Declaration of Helsinki.

Grading of images

The images were assessed by nurses and imaging technicians of non-medical background from Kitwe Central Hospital. They were trained and supervised by ophthalmologists from Frimley Park Hospital over a series of visits. These staff from Kitwe graded the retinal images at the time of photography. Each image was

graded separately for retinopathy and maculopathy using the grading system recommended for the UK National Health Service (NHS; appendix).

See Online for appendix

Subsequently, the dataset was transferred to the Singapore National Eye Centre Ocular Reading Centre (SORC) for regrading using the International Classification Diabetic Retinopathy Severity Scale (ICDRSS), for compatibility with the grading system used in the training dataset. Two SORC graders with 5 years of experience in diabetic retinopathy grading, certified and accredited for ophthalmological techniques, regraded the retinal image set into no diabetic retinopathy; mild, moderate, and severe non-proliferative diabetic retinopathy; and proliferative diabetic retinopathy. Retinal photographs with insufficient quality were flagged as ungradable if the images had more than a third of the photograph obscured. Images deemed ungradable for human graders were not included in the AI system analysis.

We defined referable diabetic retinopathy as moderate non-proliferative diabetic retinopathy or worse, including diabetic macular oedema. Ungradable images for the AI model were also flagged as referable. Vision-threatening diabetic retinopathy was defined as severe non-proliferative diabetic retinopathy and proliferative diabetic retinopathy.

Model development

We adopted an ensemble AI model consisting of a combination of two convolutional neural networks (CNNs) for classifying retinal colour fundus images: an adapted VGGNet architecture and a residual neural network (ResNet) architecture. The VGGNet architecture has been extensively used by the computer science community for extracting features from images and is characterised by a uniform structure, whereas the ResNet architecture is designed as a lower complexity structure with so-called skip residual connections.

These CNNs were trained and validated to obtain the ensemble model score (figure 1). Training of both networks was achieved by presenting the network with batches of 32 labelled images from the training dataset, incrementally exposing the network to the key characteristics of the images belonging to each class. This process enables the CNNs to gradually adapt their weight parameters to differentiate between classes through online backpropagation of errors using gradient descent, to select good invariant features. Both networks worked entirely on images and diabetic retinopathy grades, with no access to information on patient demographics or risk factors.

For the VGGNet architecture, the referable status is determined considering a multiclass classification of the 5-point ICDRSS grade, where the indices from 0 to 4 of the output nodes represent increasing diabetic retinopathy severity: a score of 0 refers to no diabetic retinopathy; scores of 1, 2, and 3 refer to mild, moderate, and severe non-proliferative retinopathy, respectively;

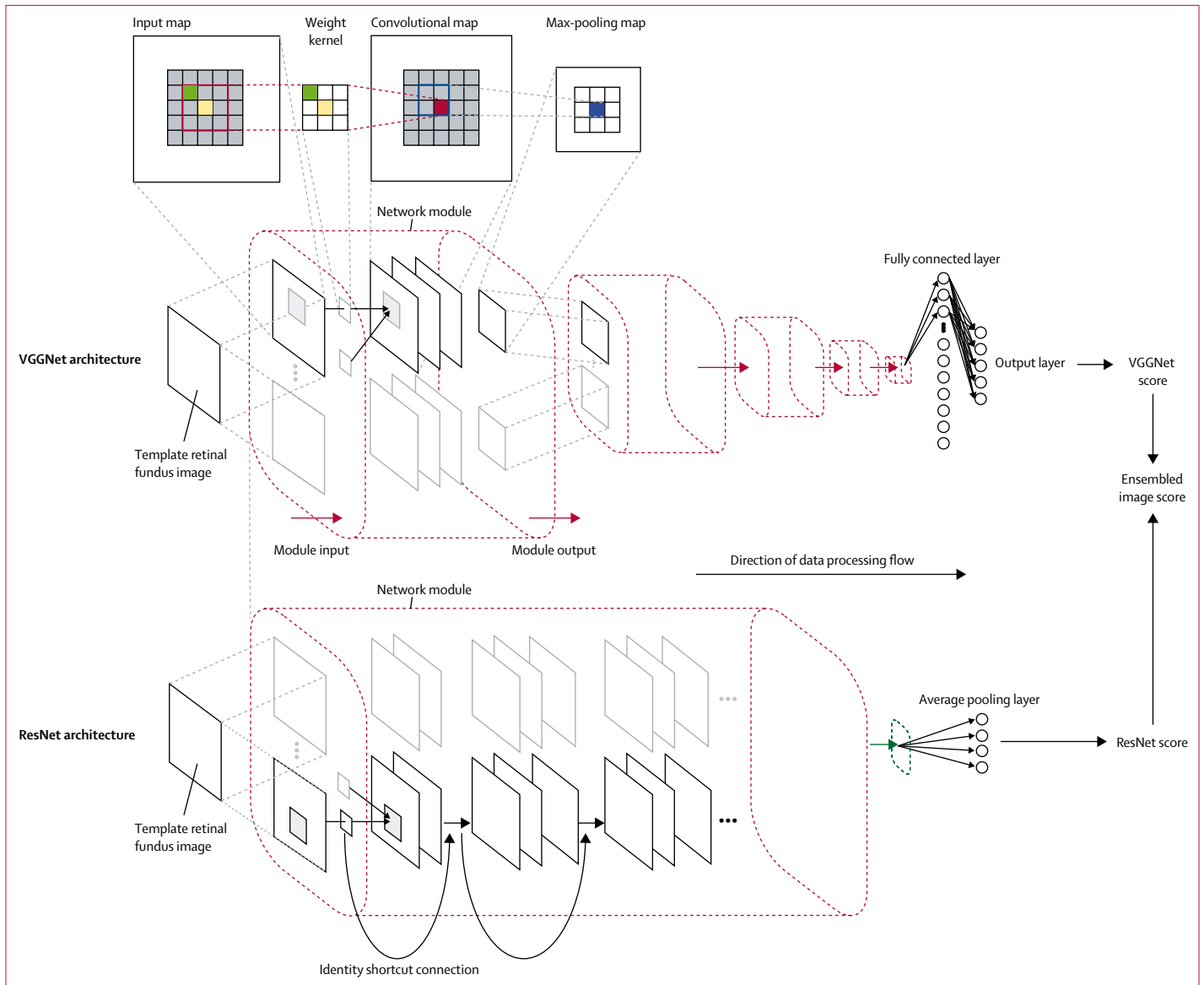


Figure 1: CNN architectures: VGGNet and ResNet models

As a preliminary step, each retinal fundus image is rescaled to fit a standardised square template of dimension 512×512 pixels. The first layer of both CNNs is then fed with the RGB values of the template image. Each VGGNet network module consists of 3×3 kernel-sized filters stacked on top of each other, with a max-pooling layer at the end of each module. The VGGNet convolutional layers are followed by a fully connected layer and terminate in a softmax classifier. Max-pooling layers perform downsampling along spatial dimensions, while the softmax classifier allows the final outputs to be interpreted as probability values by constraining their sum to 1. In total, there are 19 layers in the VGGNet architecture. By contrast, the ResNet model supports 152 layers that largely consist of numerous three-layer modules in sequence, with a final average pooling layer that computes the average values of each region as features to the final classifier. The outputs of the two CNNs are combined to create the ensemble image score. CNN=convolutional neural networks. RGB=red, green, and blue.

and a score of 4 refers to proliferative diabetic retinopathy. The final VGGNet score for referable diabetic retinopathy classification is obtained as the weighted sum of scores

$$\sum_{i=0}^4 i \cdot \text{score}(i)$$

where $\text{score}(i)$ is the probability of each output node for index i .

Unlike sequential network architectures such as VGGNet, ResNet relies on network-in-network

architectures, which are micro-architecture modules (figure 1). Residual networks allow training of deep networks by constructing the network through skip connections and feature heavy batch normalisation, both of which speed up model convergence. With this technique, the model can support more layers—ie, measure and consider more features in the images—while maintaining a lower network complexity than VGGNet. ResNet works as a multiple binary classifier and gives as outputs probabilities corresponding to four disease classifications:

	Training set (Singapore)	Validation set (Zambia)
Dataset characteristics		
Number of participants	13 099	1574
Number of eyes examined*	38 185	3093
Number of images	76 370	4504
Race or ethnicity	Chinese (9615 [73.4%]), Malay (1582 [12.1%]), Indian (1427 [10.9%]), other (407 [3.1%]), unknown (68 [0.5%])	African (1574 [100%])
Cohort	Community based	Population based
Camera	Topcon	CentreVue
Photograph technicians	3–6 months of training before certification and annual reaccreditation	3–6 months of training before certification and annual reaccreditation
Assessors	Two professional senior graders; arbitration by one retinal specialist (SORC)	Two professional senior graders (SORC) to generate the final diabetic retinopathy grading†
Patient demographics		
Mean age, years	62.8 (11.32)	55.0 (11.1)
Sex		
Female	6581 (50.2%)	885 (56.2%)
Male	6518 (49.8%)	689 (43.8%)
Smokers	..	13 (0.8%)
Systemic risk factors		
Body-mass index, mean (SD)	26.5 (4.7)	27.8 (5.9)
Median diabetes duration, years	6.4 (1.6–8.7)	4 (2–9)
Blood pressure, mm Hg		
Systolic	129.9 (16.9)	146.7 (30.5)
Diastolic	70.5 (10.1)	88.0 (13.6)
Diabetes type		
Type 1	..	121 (7.7%)
Type 2	..	357 (22.7%)
Unknown	..	1096 (69.6%)
HbA _{1c} , %	7.2 (1.4)	10.6 (5.2)
Eye characteristics‡		
Non-referable diabetic retinopathy§	36 109 (94.6%)	2396 (77.5%)
No diabetic retinopathy	33 709 (88.3%)	2318 (74.9%)
Mild non-proliferative diabetic retinopathy	3310 (8.7%)	78 (2.5%)
Referable diabetic retinopathy§	2055 (5.4%)	697 (22.5%)
Moderate non-proliferative diabetic retinopathy	597 (1.6%)	526 (17.0%)
Severe non-proliferative diabetic retinopathy	478 (1.3%)	118 (3.8%)
Proliferative diabetic retinopathy	70 (0.2%)	53 (1.7%)
Diabetic macular oedema	2026 (5.3%)	249 (8.1%)
Ungradable	21 (0.1%)	9 (0.3%)

Data are n (%) or mean (SD). SORC=Singapore Ocular Reading Centre. HbA_{1c}=glycated haemoglobin. *Patients could have more than one visit (eg, one per year) between 2010 and 2013, which means more than two eyes examined could correspond to the same patient. †The initial grading was done by a Zambian ophthalmologist and a UK professional grader. ‡Percentages are calculated on number of eyes. §Subcategories do not add up to total because eyes with diabetic retinopathy of any severity could also have diabetic macular oedema, which would render them referable.

Table 1: Summary of the training and validation datasets

mild non-proliferative diabetic retinopathy or worse (p_0), moderate non-proliferative diabetic retinopathy or worse (p_1), severe non-proliferative diabetic retinopathy or worse (p_2), and proliferative diabetic retinopathy (p_3). The final score for referable diabetic retinopathy classification is p_1 —ie, of any type of diabetic retinopathy excluding mild.

The ensemble model is finally built as a combination of these VGGNet and ResNet probability output scores,

summing the probabilities given by each model of referable diabetic retinopathy:

$$p_1 + \sum_{i=0}^4 i \cdot \text{score}(i)$$

The final classification is obtained by thresholding the algorithms' output scores for the desired sensitivity and specificity performance (figure 1); for predetermined

	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Referable diabetic retinopathy*	0.973 (0.969–0.978)	92.25% (90.10–94.12)	89.04% (87.85–90.28)
Vision-threatening diabetic retinopathy†	0.934 (0.924–0.944)	99.42% (99.15–99.68)	..
Diabetic macular oedema	0.942 (0.933–0.952)	97.19% (96.61–97.77)	..

Eyes were the units of analysis (n=3093). The model was designed to identify referable diabetic retinopathy, hence specificities are not reported for vision-threatening diabetic retinopathy and diabetic macular oedema. AUC=area under the curve. *Defined as moderate non-proliferative, severe non-proliferative, or proliferative diabetic retinopathy; diabetic macular oedema; and ungradable eyes. †Defined as severe non-proliferative or proliferative diabetic retinopathy.

Table 2: Referable diabetic retinopathy AUC, sensitivity, and specificity of the ensemble artificial intelligence model in the validation dataset and vision-threatening diabetic retinopathy and diabetic macular oedema AUCs and sensitivities among referable eyes

optimal sensitivity of at least 90%—suitable for the screening use case examined here—these thresholds were determined as 0.70 for VGGNet, 0.13 for ResNet, and 0.43 for the ensemble.

Model training

The model was trained with 76 370 retinal fundus images from 13 099 patients with diabetes who participated in the Singapore Integrated Diabetic Retinopathy Program, between 2010 and 2013. The fundus images were graded according to the ICDRSS by two professional senior graders (more than 5 years' experience) with the arbitration of a retinal specialist (PhD-trained with more than 5 years' experience in diabetic retinopathy assessment). The same dataset was used by Ting and colleagues²² to train their AI system, based on a VGGNet model only. In this study, we trained an additional model (ResNet) and consider an ensemble of the two.

To assess the relationship between diabetic retinopathy and potential risk factors, we considered a patient-level perspective. For each patient, we considered the grade of the eye with the most severe diabetic retinopathy level as predicted. In the case in which one of the two eyes resulted ungradable, the grade of the other eye was used. If both eyes were ungradable, the patient was assigned as ungradable and excluded from the analysis.

To understand the behaviour of the AI system in making decisions, we generated heat maps that highlighted the visual features of the processed image that contributed the most to the output assignment of the predicted diagnoses or led to unappropriated misclassification. We used explanations from integrated gradient,³¹ a gradient-based method that generates pixel-based masks and measures the contribution of each pixel with respect to a baseline image.

Statistical analysis

Our hypothesis is that the AI model is at least comparable to the human graders' performance. The operating threshold was pre-set at 90% sensitivity on the basis of our previous study,²² using the criteria set by the Ministry of Health of Singapore. We calculated the area under the curve (AUC) for the receiver operating characteristic

curve, sensitivity, and specificity for AI system in the detection of referable diabetic retinopathy. Among the referable diabetic retinopathy cases, the sensitivity (ie, detection rate) of vision-threatening diabetic retinopathy and diabetic macular oedema were also calculated. Cluster-bootstrap biased-corrected 95% CIs were computed for each AUC, where individual patients were the bootstrap sampling clusters. Asymptotic two-sided 95% CIs were computed for the logit transform of each proportion (ie, sensitivity and specificity). The cluster sandwich estimator of the SE was used to account for possible dependency of eyes within each individual.

We then did a multivariable logistic regression on the risk factors for AI model and human grader outcomes in detecting referable diabetic retinopathy. The association of diabetic retinopathy with risk factors was assessed using odds ratios and was compared between AI system and human graders for statistical difference, using the two-tailed Student's *t*-test. Finally, we calculated the AUC of the model to evaluate the discriminative ability of the combined risk factors.

We used a significance level of $\alpha=0.05$ for our analyses. All analyses were done using R (version 3.5.1).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

A total of 4504 images (1574 patients, 3093 eyes) from the validation dataset were analysed in this study (table 1). Dataset characteristics and patient demographics, diabetes history, and systemic risk factors from both datasets are summarised in table 1. The median duration of diabetes was 4 years (IQR 2–9), and diabetes type was known for less than a third of participants (table 1). Referable diabetic retinopathy was found in 697 (22.5%) eyes, vision-threatening diabetic retinopathy in 171 (5.5%) eyes, and diabetic macular oedema in 249 (8.1%) eyes; there were nine (0.3%) ungradable eyes in the dataset (table 1).

The AUC of the AI system for referable diabetic retinopathy was 0.973 (95% CI 0.969–0.978), with 92.25% (90.10–94.12) sensitivity and 89.04% (87.85–90.28) specificity (table 2, figure 2). Of the referable eyes, the sensitivity was 99.42% (99.15–99.68) for vision-threatening diabetic retinopathy and 97.19% (96.61–97.77) for diabetic macular oedema, with AUCs of 0.934 (0.924–0.944) and 0.942 (0.933–0.952), respectively (table 2, figure 2). The ensemble method produced comparable performance in different subgroups of patients stratified by age, sex, and HbA_{1c} (appendix).

We used heat maps to highlight the area of the features in the retinal fundus images that contributed most to the CNN's assignment of the predicted diagnoses (figure 3).

These visualisations indicate how the AI system makes decisions and represent feature importance explanations that could provide a rationale to build trust in AI models.

In multivariable analysis, we found that longer duration of diabetes, increased HbA_{1c}, and higher systolic blood pressure were significantly associated with referable diabetic retinopathy for both the AI model and human assessors (table 3). Female sex was found to be a predictor for diabetic retinopathy when detected by human graders but not by the AI model. Systemic risk factors were comparable between AI model and human graders in detecting referable diabetic retinopathy: the AUC was 0·723 (95% CI 0·691–0·754) for the AI model and 0·741 (0·710–0·771) for human graders ($p=0·432$; appendix).

Discussion

In this study, we showed that the AI model is clinically effective in detecting referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema within a real-life diabetic retinopathy screening programme in Zambia, where the retinal images were mainly captured by non-medically trained technicians. Although most AI studies have been done in high-income regions such as the USA, Europe, China, and Singapore, it is important to evaluate the application of this sophisticated technology in detection of diabetic retinopathy in low-resource countries. Our findings suggest that the application of AI to analysis of retinal images could provide an alternative solution for diabetic retinopathy screening, especially in settings with little access to human expertise. Furthermore, we compared the AI model and human graders in referable diabetic retinopathy prevalence detection and systemic risk factors associations, showing similar outcomes in both the tasks. Specifically, both the AI model and human graders identified longer duration of diabetes, higher level of HbA_{1c}, and increased systolic blood pressure as risk factors associated with diabetic retinopathy.

Before the deep learning era, Hansen et al³² reported an AI system that used feature-based learning in detection of any diabetic retinopathy, showing an AUC of 0·878 with a negative predictive value of 98%. In our study, the ensemble AI model was shown to have excellent AUC, sensitivity, and specificity in detecting referable diabetic retinopathy in an African population, even when the algorithm was trained using different ethnic groups (Chinese, Malay and Indian), and different retinal cameras, image resolution, and width of field. Similarly, in the previous study,²² this AI system also showed good testing results on African American Eye Disease Study datasets (0·983 AUC, 98·8% sensitivity, and 82·0% specificity), suggesting the consistency and generalisability of this AI system in detecting referable diabetic retinopathy for patients with dark fundi.

Among the referable cases of diabetic retinopathy, we computed the detection rate for vision-threatening diabetic retinopathy and diabetic macular oedema,

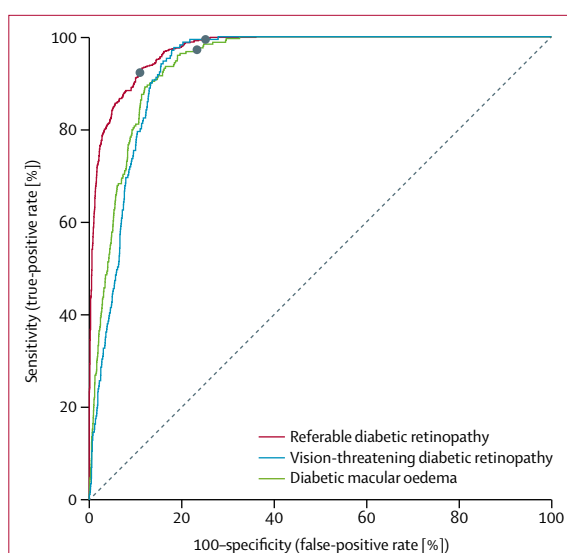


Figure 2: Receiver operating characteristic curve of the ensemble artificial intelligence model for detection of referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema in the validation dataset

Eyes are the units of analysis ($n=3093$). Points on the curve indicate the final sensitivities (and specificity, for referable diabetic retinopathy) of the model according to the desired thresholds (see table 2). As per definition, vision-threatening diabetic retinopathy is severe non-proliferative or proliferative diabetic retinopathy condition. The receiver operating characteristic curve of vision-threatening diabetic retinopathy considered a prediction of mild or moderate diabetic retinopathy as a false negative.

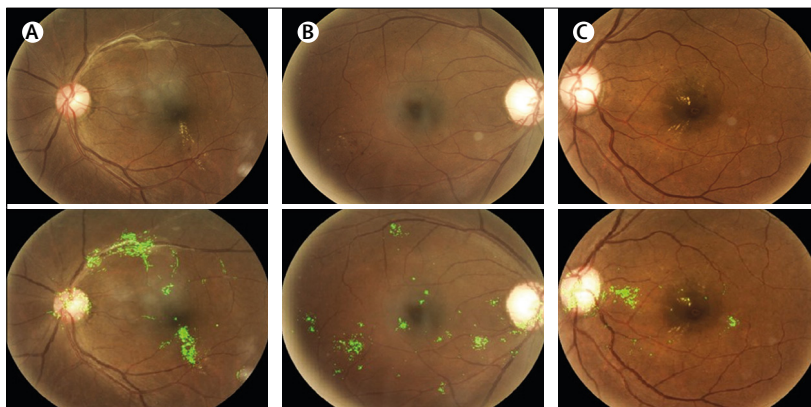


Figure 3: Heat map visualisations using integrated gradient method

In these positive cases of referable diabetic retinopathy, the green coloured areas in the retinal fundus images show the contribution to the artificial intelligence model's assignment of the predicted referable diabetic retinopathy diagnosis. The images show fibrovascular proliferation associated with hard exudate located close to fovea region, suggesting possibility of diabetic macular oedema (A); moderate diabetic retinopathy with subtle changes in the area around macula as well as inferior temporal quadrant (B); and hard exudates and haemorrhages located close to the fovea region, suggesting diabetic macular oedema (C). The modifications flagged in green are sometimes missed by retinal specialists and ophthalmologists due to poor image quality and heat map visualisation can aid in making a diagnosis. The resolution of the fundus overlay (512 × 512 pixels) produced by the model is slightly smaller than the original image resolution, due to the down-sampling effect during the pre-processing phase.

obtaining sensitivities of 99·42% and 97·19%, respectively, showing that critical positive cases were very rarely missed. Specifically, given that high prevalence of vision-threatening diabetic retinopathy is a surrogate marker of blindness, this study shows how AI can aid early diagnosis

	Artificial intelligence model		Human graders		Group difference	
	OR (95% CI)	p value (logistic regression)	OR (95% CI)	p value (logistic regression)	Difference in standardised ORs	p value*
Age (per 1-year increase)	1.06 (0.92 to 1.24)	0.42	0.96 (0.83 to 1.13)	0.65	0.10 (−0.12 to 0.31)	0.37
Female sex (vs male)	0.77 (0.58 to 1.02)	0.067	0.71 (0.53 to 0.95)	0.023	0.06 (−0.35 to 0.46)	0.72
Body-mass index	0.88 (0.76 to 1.02)	0.088	0.95 (0.82 to 1.09)	0.46	0.06 (−0.14 to 0.27)	0.50
Type 1 diabetes (vs type 2)	1.16 (1.03 to 1.31)	0.013	1.07 (0.94 to 1.20)	0.32	0.10 (−0.08 to 0.27)	0.32
Diabetes duration (per 1-year increase)	1.77 (1.54 to 2.04)	<0.0001	1.87 (1.63 to 2.16)	<0.0001	0.10 (−0.10 to 0.30)	0.58
HbA _{1c}	1.21 (1.06 to 1.38)	0.0035	1.30 (1.14 to 1.48)	0.0001	0.08 (−0.1 to 0.27)	0.47
Systolic blood pressure	1.51 (1.25 to 1.85)	<0.0001	1.44 (1.18 to 1.76)	0.00039	0.08 (−0.2 to 0.36)	0.72
Diastolic blood pressure	0.94 (0.79 to 1.12)	0.47	1.07 (0.90 to 1.29)	0.44	0.14 (−0.12 to 0.39)	0.29

Patients were the units of analysis (n=1574). OR=odds ratio. HbA_{1c}=glycated haemoglobin. *p value for the difference in standardised ORs in the multivariable regression between deep learning system and human assessors, generated using a two-tailed Student's t-test.

Table 3: Multivariable analysis of systemic risk factors with referable diabetic retinopathy diagnosed by the ensemble artificial intelligence model, as compared with human graders

and enable the small number of ophthalmologists in these regions to focus on the treatment of vision-threatening diabetic retinopathy.

When exploring and regrading false-positive and false-negative images, some discrepancies were found between Kitwe Central Hospital–Frimley Park Hospital grades and SORC grades, the former using the UK NHS grading scale and the latter deploying the ICDRSS. The greatest discrepancy between the scales in how the various conditions are classified are in the classification of diabetic macular oedema (appendix). Standardisation of diabetic retinopathy classifications is therefore needed to facilitate the AI deployment to the global population with diabetes. Moreover, the definition of referable diabetic retinopathy in a resource-rich country versus a low-resource country might be entirely different. In the high-income countries with well established diabetic retinopathy screening programmes, the AI system will more often be trained to err on the side of caution by lowering the threshold of referrals, but with understanding that the over-referral of false positives might not create a workforce issue in the tertiary setting. However, over-referrals of relatively stable patients with diabetes (eg, with moderate non-proliferative diabetic retinopathy) to consult ophthalmologists in the tertiary care setting in Zambia could create some challenges. Thus, the referable diabetic retinopathy for this population might be redefined as severe non-proliferative and proliferative diabetic retinopathy and diabetic macular oedema.

Although the AI system was shown to be clinically acceptable to detect referable diabetic retinopathy in this African population, a future study testing this algorithm on more Zambian patients—eg, patients with milder disease, patients with shorter duration of diabetes, or untreated cases—would be of great value to ensure a wider generalisability of this algorithm. Additionally, studies using other portable retinal cameras with different image characteristics or in other African populations are encouraged. Such studies might also

allow the algorithm to be fine-tuned to detect more granular disease severity levels.

Our study has several strengths. First, this is one of the few AI studies to date that show the use of AI using deep learning in under-resourced countries with no existing diabetic retinopathy screening programme in place. Second, this AI system showed excellent generalisability for patients from different ethnicities in detection of referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema. Third, the retinal images used were obtained in a prospective manner within real-world clinical settings.

Our study also has several weaknesses. First, the study population was limited to patients with self-reported diabetes of whom almost 70% did not know their diabetes type. Second, the affordability and usability of the AI system considered is questionable for under-resourced countries such as Zambia, especially if they have a suboptimal telecommunication network. Although the combined ensemble model generated the most accurate results when compared with VGGNet and ResNet individually, this combined model also requires more computational power and running time, as well as telecommunication support if it were to be deployed as a cloud-based model. To clinically deploy and ensure successful adoption of a sophisticated AI system in a low-income country, it is important to consider these factors. Another potentially useful strategy could be to incorporate the AI system into the retinal cameras or use AI as a standalone system that does not require the sophisticated telecommunication network. These strategies could potentially increase the clinical adoption rate in the under-resourced countries worldwide. Third, while our study shows a reliable method for diabetic retinopathy screening in under-resourced countries, once patients have been diagnosed with referable or vision-threatening diabetic retinopathy, it remains uncertain how many of these patients will eventually receive treatment to avoid diabetic retinopathy-related blindness. Therefore, our screening

method needs to be adopted alongside an effective treatment strategy to have its intended impact.

In conclusion, our study shows a clinically acceptable AI system in detection of referable diabetic retinopathy, vision-threatening diabetic retinopathy, and diabetic macular oedema for the Zambia population. Future research is needed to evaluate the cost-effectiveness of such sophisticated technology for diabetic retinopathy screening worldwide, especially for those countries with little access to health-care services. The joint UK–Zambian initiative, coupled with AI application, could potentially lead to a novel model of care for diabetic retinopathy screening. It is, however, also important to ensure that once detected with vision-threatening diabetic retinopathy, these patients can also be seen and treated promptly in the tertiary eye-care services in Zambia. With a comprehensive screening and referral system, this would help to improve access to and quality of eye-care services in Zambia, aligning with the VISION 2020 strategy to eliminate preventable blindness. Nevertheless, whether or not this approach would liberate highly demanded ophthalmologists from the task of interpreting images and allow them to focus instead on treating vision-threatening diabetic retinopathy or diabetic macular oedema still remains to be answered. Compared with resource-rich countries, AI application within low-resource settings might need more careful and comprehensive design and planning, taking into account the availability of the specialists, long-term patient outcomes, and the cost-effectiveness of the AI screening programme.

Contributors

VB, SS, TYW, GM, and DSWT contributed to the initial drafting of the manuscript. LM, MC, GM, MM, SS, and GC-M contributed to the data collection. VB, ZWL, GL, QDN, YX, MYTY, HH, JH, XQL, WH, and MLL contributed to the data analysis. All authors contributed to the critical review and final approval for this manuscript.

Declaration of interests

DSWT, GL, MLL, WH, and TYW are co-inventors of a patent on the deep learning system in this paper; potential conflicts of interests are managed according to institutional policies of the Singapore Health System (SingHealth) and the National University of Singapore. All other authors declare no competing interests.

Acknowledgments

This study was funded by the National Medical Research Council Health Service Research Grant, Large Collaborative Grant, Singapore Ministry of Health (National Health Innovation Center, Innovation to Develop Grant NHIC-12D-1409022); SingHealth Foundation Research Grant (SHF/FG648S/2015); and the Tanoto Foundation. The Singapore Diabetic Retinopathy Program received funding from the Singapore Ministry of Health (grants AIC/RPDD/SIDRP/SERI/FY2013/0018 and AIC/HPD/FY2016/0912). The Diabetes Study in Nephropathy and Other Microvascular Complications received funding from the National Medical Research Council Large Collaborative Grant.

References

- 1 International Monetary Fund. World economic outlook database. October 2018. Washington, DC: International Monetary Fund, 2018.
- 2 Doctor HV, Nkhana-Salimu S, Abdulsalam-Anibilowo M. Health facility delivery in sub-Saharan Africa: successes, challenges, and implications for the 2030 development agenda. *BMC Public Health* 2018; **18**: 765.
- 3 Peer N, Kengne AP, Motala AA, Mbanya JC. Diabetes in the Africa Region: an update. *Diabetes Res Clin Pract* 2014; **103**: 197–205.
- 4 Hall V, Thomsen RW, Henriksen O, Lohse N. Diabetes in Sub Saharan Africa 1999–2011: epidemiology and public health implications. A systematic review. *BMC Public Health* 2011; **11**: 564.
- 5 Sivaprasad S, Gupta B, Crosby-Nwaobi R, Evans J. Prevalence of diabetic retinopathy in various ethnic groups: a worldwide perspective. *Surv Ophthalmol* 2012; **57**: 347–70.
- 6 Burgess PI, MacCormick IJ, Harding SP, Bastawrous A, Beare NA, Garner P. Epidemiology of diabetic retinopathy and maculopathy in Africa: a systematic review. *Diabet Med* 2013; **30**: 399–412.
- 7 Resnikoff S, Felch W, Gauthier T-M, Spivey B. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners. *Br J Ophthalmol* 2012; **96**: 783–87.
- 8 Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: the International Council of Ophthalmology Recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology* 2018; **125**: 1608–22.
- 9 Ting DSW, Cheung GCM, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. *Clin Exp Ophthalmol* 2016; **44**: 260–77.
- 10 Flaxman SR, Bourne RR, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health* 2017; **5**: e1221–34.
- 11 Nangia V, Jonas JB, George R, et al. Prevalence and causes of blindness and vision impairment: magnitude, temporal trends and projections in South and Central Asia. *Br J Ophthalmol* 2018; published online Nov 8. DOI:10.1136/bjophthalmol-2018-312292.
- 12 Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012; **35**: 556–64.
- 13 Burgess PI, Harding SP, García-Fiñana M, Beare NA, Msukwa G, Allain TJ. First prospective cohort study of diabetic retinopathy from sub-Saharan Africa: high incidence and progression of retinopathy and relationship to human immunodeficiency virus infection. *Ophthalmology* 2016; **123**: 1919–25.
- 14 Mathenge W, Bastawrous A, Peto T, et al. Prevalence and correlates of diabetic retinopathy in a population-based survey of older people in Nakuru, Kenya. *Ophthalmic Epidemiol* 2014; **21**: 169–77.
- 15 Bastawrous A, Mathenge W, Wing K, et al. The incidence of diabetes mellitus and diabetic retinopathy in a population-based cohort study of people age 50 years and over in Nakuru, Kenya. *BMC Endocr Disord* 2017; **17**: 19.
- 16 Burgess PI, Harding SP, García-Fiñana M, et al. Incidence and progression of diabetic retinopathy in Sub-Saharan Africa: a five year cohort study. *PloS One* 2017; **12**: e0181359.
- 17 Lewis AD, Hogg RE, Chandran M, et al. Prevalence of diabetic retinopathy and visual impairment in patients with diabetes mellitus in Zambia through the implementation of a mobile diabetic retinopathy screening project in the Copperbelt province: a cross-sectional study. *Eye (Lond)* 2018; **32**: 1201–08.
- 18 Ting DS, Yi PH, Hui F. Clinical applicability of deep learning system in detecting tuberculosis with chest radiography. *Radiology* 2018; **286**: 729.
- 19 Bejnordi BE, Veta M, Van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318**: 2199–210.
- 20 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316**: 2402–10.
- 21 Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci* 2016; **57**: 5200–06.
- 22 Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017; **318**: 2211–23.
- 23 Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018; **1**: 39.
- 24 Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care* 2018; **41**: 2509–16.

- 25 Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017; **124**: 962–69.
- 26 Hood DC, De CM. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018; **125**: 1199–206.
- 27 Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol* 2017; **135**: 1170–76.
- 28 Burlina PM, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. *JAMA Ophthalmol* 2018; **136**: 1359–66.
- 29 Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol* 2018; **136**: 803–10.
- 30 Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomedl Eng* 2018; **2**: 158.
- 31 Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. *Proc Mach Learn Res* 2017; **70**: 3319–28.
- 32 Hansen MB, Abramoff MD, Folk JC, Mathenge W, Bastawrous A, Peto T. Results of automated retinal image analysis for detection of diabetic retinopathy from the Nakuru study, Kenya. *PLoS One* 2015; **10**: e0139148.